

7. En Weka grafique el árbol de decisión del dataset seleccionado.

Debido a que con el dataset que manejamos tenemos valores que pueden ser representados con $f(x)$ de varias formas nos centraremos en escoger una clase para realizar el árbol de decisión, para lo cual eliminaremos atributos o columnas que podemos obviar en nuestro análisis, de esta forma limpiaremos los datos para dejar los más relevantes

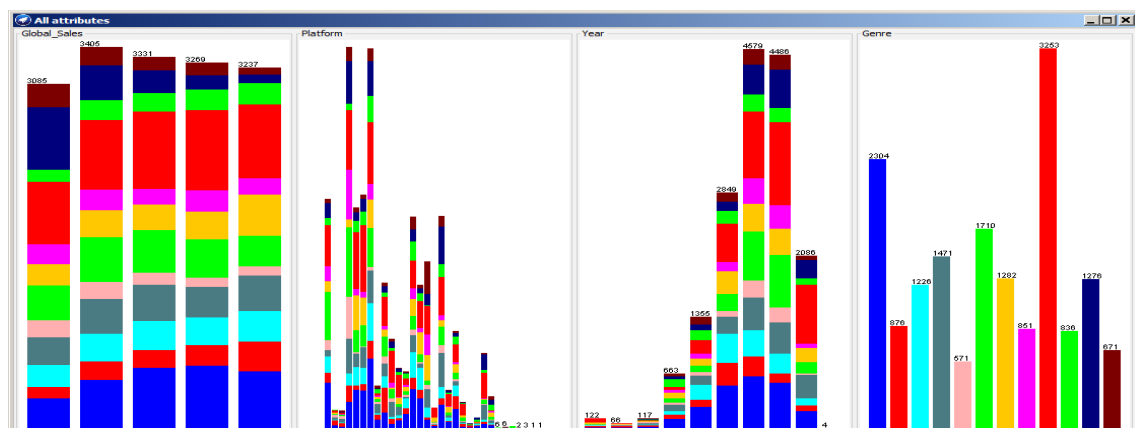
Globlal_Sales	Nos indica el valor de las ventas globales de un determinado juego de un género específico (es la suma total por lo tanto el resumen de las ventas de los atributo eliminados)
Platform	Representa la consola o plataforma para la cual fue lanzada un juego
Year	Año de salida de los juegos
Genre	(clase) Género al cual pertenece el video juego

Nota: Cabe mencionar que en los casos de valores perdidos son obviados puesto que valores perdidos, no son utilizados en el análisis del árbol de decisión pero si podrían ser usados para casos de testeo.

Preprocesamientos

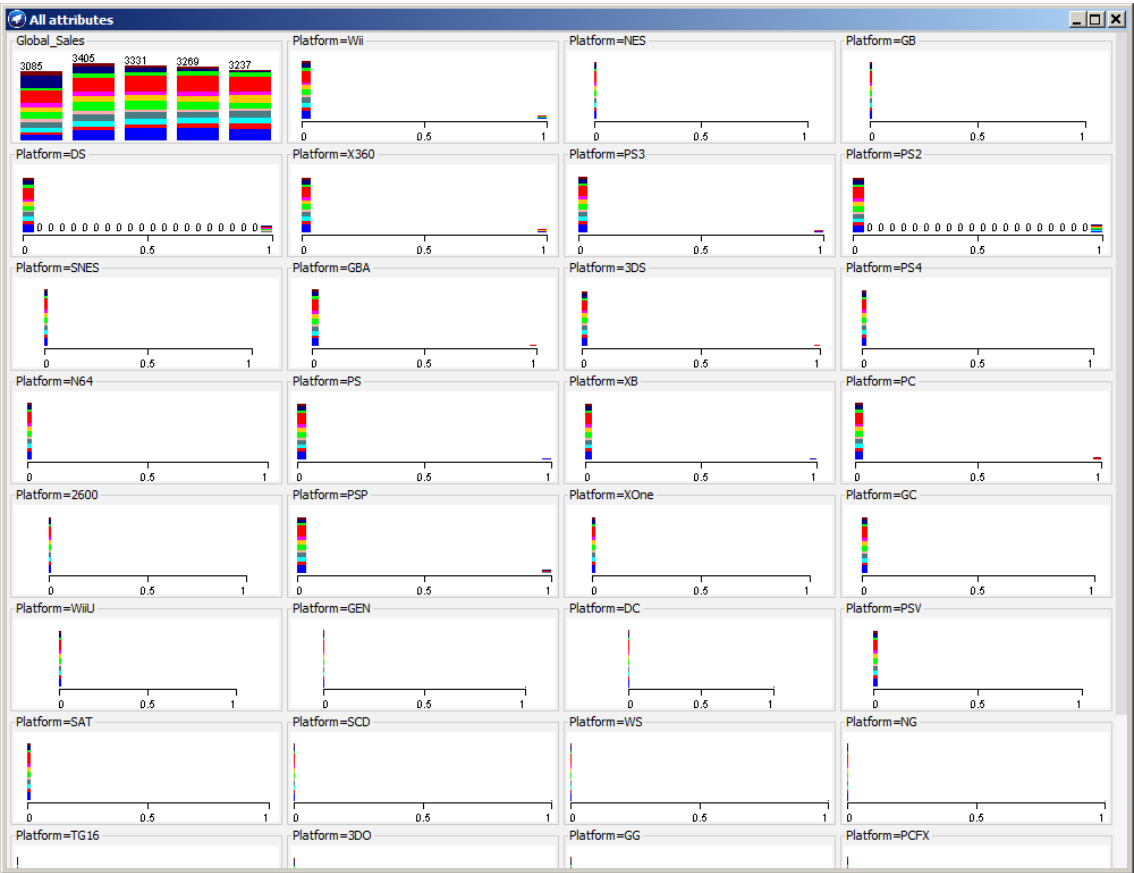
Globlal_Sales	Utilizaremos discretización (5 rangos)
Platform	Utilizaremos el onehot (NominalToBinary de weka) lo cual añadirá varios atributos o columnas al dataset
Year	Utilizaremos discretización (10 rangos)
Genre	No cambiara

Imágenes de los atributos discretizados



Imágenes de los atributos con onehot

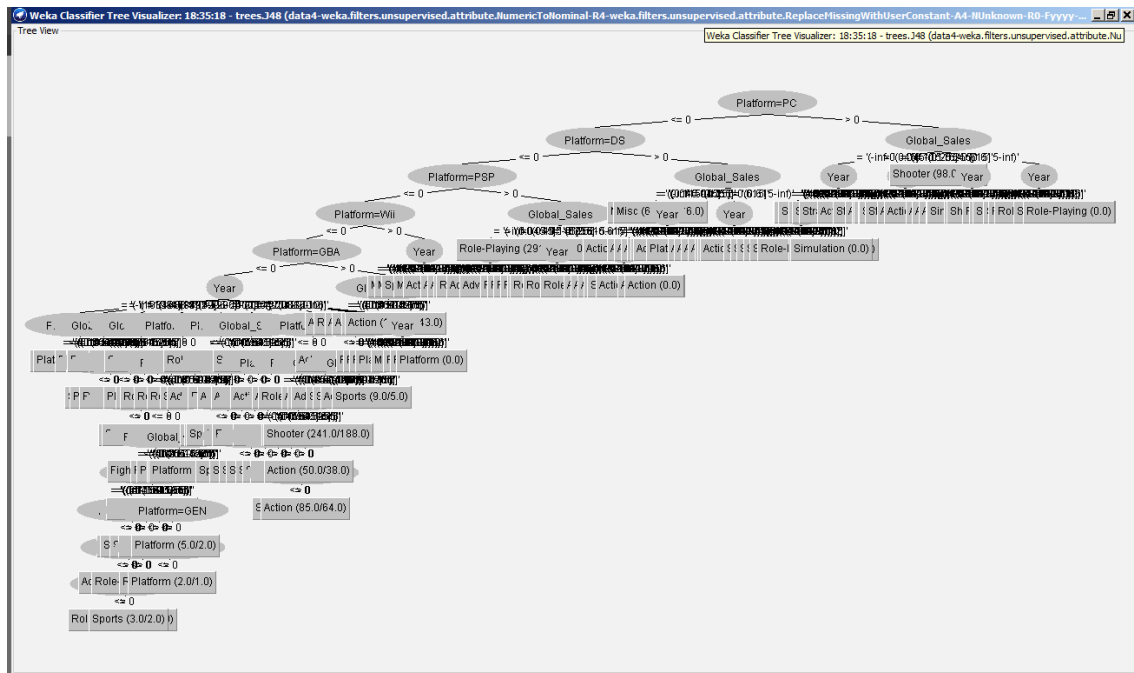
Nombre: Landaeta Navarro Adolfo Antonio



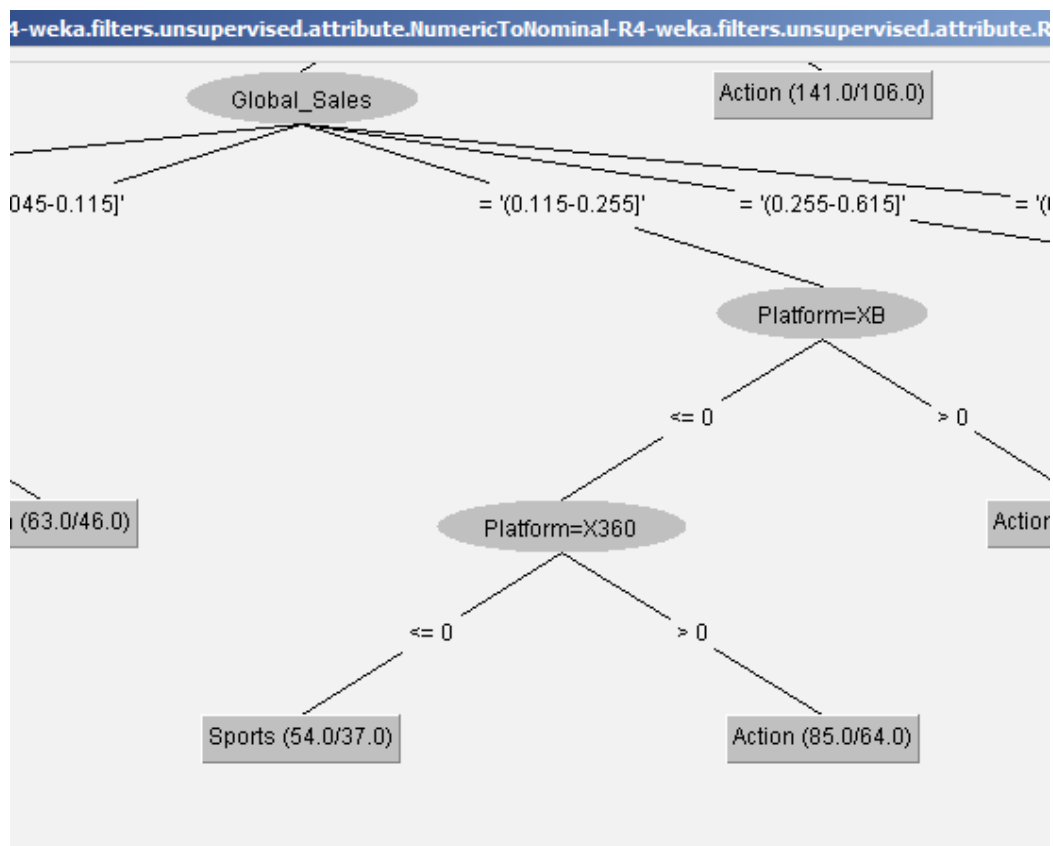
Generamos el árbol de decisión con estos 4 atributos, usando el algoritmo J48y este es nuestro resultado

Classifier output									
=== Stratified cross-validation ===									
=== Summary ===									
Correctly Classified Instances	3921			24.0154 %					
Incorrectly Classified Instances	12406			75.9846 %					
Kappa statistic	0.1117								
Mean absolute error	0.143								
Root mean squared error	0.2687								
Relative absolute error	96.1585 %								
Root relative squared error	98.5502 %								
Total Number of Instances	16327								
=== Detailed Accuracy By Class ===									
TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class	
0,365	0,238	0,201	0,365	0,259	0,101	0,628	0,194	Sports	
0,074	0,011	0,273	0,074	0,117	0,118	0,666	0,118	Platform	
0,033	0,009	0,240	0,033	0,059	0,064	0,646	0,122	Racing	
0,099	0,045	0,179	0,099	0,128	0,071	0,625	0,136	Role-Playing	
0,018	0,003	0,172	0,018	0,032	0,045	0,750	0,096	Puzzle	
0,288	0,133	0,201	0,288	0,237	0,132	0,656	0,170	Misc	
0,069	0,026	0,186	0,069	0,100	0,069	0,673	0,133	Shooter	
0,053	0,012	0,195	0,053	0,083	0,077	0,681	0,111	Simulation	
0,550	0,358	0,277	0,550	0,368	0,157	0,628	0,283	Action	
0,036	0,006	0,246	0,036	0,063	0,077	0,651	0,094	Fighting	
0,203	0,025	0,409	0,203	0,271	0,247	0,701	0,225	Adventure	
0,173	0,024	0,233	0,173	0,198	0,171	0,691	0,117	Strategy	
Weighted Avg.	0,240	0,130	0,238	0,240	0,203	0,654	0,176		

Vemos que el error es demasiado grande por lo cual deberíamos mejorar los datos que tenemos, posiblemente tener más cantidad u otra clasificación para mejorar el grado de detección del algoritmo, si analizamos el árbol con estos datos veremos.



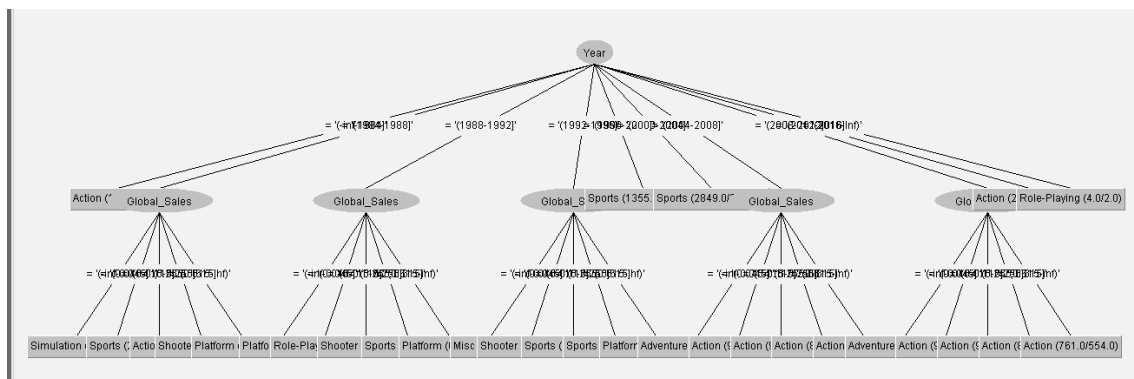
El árbol puede representar varios valores de decisión, tomando como ejemplo podemos ver que en cuanto a la plataforma Xbox360 los juegos de tipo “sport” y “action” son bastante solicitados en lo que va a ser los años 2004-2008.



Nombre: Landaeta Navarro Adolfo Antonio

Comprobando el dataset y eliminando el atributo de plataforma volvemos un nuevo árbol de decisión, que vemos que mejora la predicción, lo cual nos ayuda a interpretar que quizá el atributo de plataforma es un atributo demasiado complejo lo cual deberíamos buscar una nueva forma de clasificarlo

Classifier output		
Number of Leaves :	30	
Size of the tree :	36	
Time taken to build model: 0.01 seconds		
=== Stratified cross-validation ===		
=== Summary ===		
Correctly Classified Instances	3570	21.8656 %
Incorrectly Classified Instances	12757	78.1344 %
Kappa statistic	0.0643	
Mean absolute error	0.1461	
Root mean squared error	0.2705	
Relative absolute error	98.2371 %	
Root relative squared error	99.2162 %	
Total Number of Instances	16327	



Analizando el árbol vemos que en el intervalo de los años 2012-2016 los juegos de acción tienen un valor muy elevado por lo cual podemos decir que este tipo de juegos era bastante solicitado, y en cuanto a ventas en intervalos 2008-2012 los juegos de acción y aventura fueron los que presentaron mayor ganancia económica.

